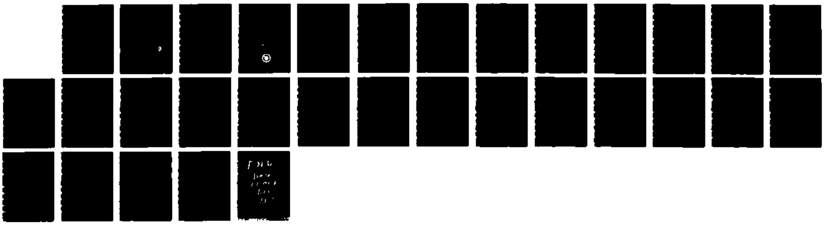
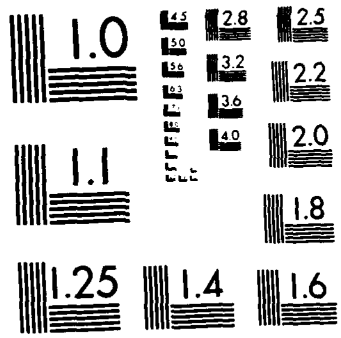
AD-A185 525 TESTING AND INTERVAL ESTIMATION IN A CHANGE-POINT MODEL 1/1
ALLOWING AT MOST (U) PITTSBURGH UNIV PA CENTER FOR
MULTIVARIATE ANALYSIS X CHEN JUL 87 TR-87-25
UNCLASSIFIED AFOSR-TR-87-0972 F49620-85-C-0008 F/G 12/3 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Unclassified

DTIC FILE COPY (2)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A185 525

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 87-0972	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Testing and interval estimation in a change-point model allowing at most one change		5. TYPE OF REPORT & PERIOD COVERED ~~Technical~~ - July 1987
		6. PERFORMING ORG. REPORT NUMBER 87-25
7. AUTHOR(s) Xiru Chen		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis, 515 Thackeray Hall, University of Pittsburgh, Pittsburgh PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research, Department of the Air Force, Bolling Air Force Base, DC 20332		12. REPORT DATE July 1987
		13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same as 11		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE OCT 0 2 1987 S D

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Brownian Motion Process
Change Point
Asymptotic Distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper considers the simplest model of change-point in which at most one change in the mean may occur. Results include:

1. Introduction of a test for the null hypothesis that "no change in the mean occurs", and the limit distribution of the test-statistic.
2. Approximate calculation of the power of the test.
3. Interval estimation of the position of change.
4. Point estimation of the jump at the point of change and its asymptotic distribution.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

87 9 24 219

5. Evaluation of the bias of the MLE of error variance.

AFOSR-TR- 87-0972

TESTING AND INTERVAL ESTIMATION IN A CHANGE-POINT MODEL ALLOWING AT MOST ONE CHANGE

Xiru Chen

Center for Multivariate Analysis
University of Pittsburgh

Center for Multivariate Analysis

University of Pittsburgh

TESTING AND INTERVAL ESTIMATION IN A CHANGE-POINT MODEL ALLOWING AT MOST ONE CHANGE

Xiru Chen

Center for Multivariate Analysis
University of Pittsburgh

July 1987

Technical Report No. 87-25

Center for Multivariate Analysis
University of Pittsburgh
515 Thackeray Hall
Pittsburgh, PA 15260

Research sponsored by the Air Force Office of Scientific Research(AFSC), under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

TESTING AND INTERVAL ESTIMATION IN A CHANGE-POINT MODEL ALLOWING AT MOST ONE CHANGE

Xiru Chen

Center for Multivariate Analysis
University of Pittsburgh

ABSTRACT

This paper considers the simplest model of change-point in which at most one change in the mean may occur. Results include:

1) Introduction of a test for the null hypothesis that "no change in the mean occurs", and the limit distribution of the test-statistic.
2) Approximate calculation of the power of the test.
3) Interval estimation of the position of change.
4) Point estimation of the jump at the point of change and its asymptotic distribution. and
5) Evaluation of the bias of the MLE of error variance.

(Keywords:

AMS 1980 subject classifications: Primary 62F03; Secondary 62F12

Key words and phrases: Brownian Motion Process), (en)
Change Point
Asymptotic Distribution

TESTING AND INTERVAL ESTIMATION IN A CHANGE-POINT MODEL ALLOWING AT MOST ONE CHANGE

Xiru Chen

1. INTRODUCTION

Consider the model

$$X(t) = f(t) + e(t), \quad 0 \le t \le 1 \tag{1}$$

where f(t) is a non-random function of the form

$$f(t) = \begin{cases} a, & 0 \le t \le t_0 \\ a+\theta, & t_0 < t \le 1 \end{cases}$$

with a, θ, t_0 unknown and $0 < t_0 < 1$. t_0 is called the "change-point" (of the function f, or of model (1)), and θ the jump at the change-point. e(t) is a random variable whose distribution function F is independent upon t. We assume that $E(e(t)) = 0$ for $t \in [0, 1]$, and e(t) possesses a finite variance σ^2.

We desire to make inference on t_0 and θ, using observations made on $\{X(t)\}$. In this paper we assume that these observations are taken on equal-paced t values. Specifically, we observe $X(i/n)$, $i = 1,\ldots,n$. Note that $X(i/n)$ depends upon both i and n, but for simplicity of notation in the sequel we write X_i instead of $X(i/n)$. We assume that $X_1,\ldots,X_n$ are independent. These assumptions and notations, such as a, θ, σ^2, F are valid throughout this paper, and will not be mentioned latter.

Model (1), the so-called AMOC (Allowing at Most One Change), is the simplest model for change-point. As an important and extensively studied model, there exists a huge literature on it, and we refer to Krishnaiah and Miao (1986), Csörgo and Horvath (1986) for a detailed survey of this subject.

The methodology of the present paper is nonparametric in nature, in that we do not resort to likelihood and thus, the normality condition can be dispensed with. The work borrows an idea advanced by Yin (1986). This idea proposes to search the possible change-point by comparisons made locally. The method has the merit that the resulted statistic in testing the null hypothesis

$$H_0 : \quad \theta = 0 \text{ (i.e. no change-point exists)} \tag{3}$$

has a simple asymptotic distribution. This not only facilitates the testing of H_0, but also is convenient in estimating the power of test, and in constructing a confidence interval for the change-point t_0. These problems will be studied in Section 2, Section 3 and Section 5, with different assumptions on F. Section 6 is devoted to a brief discussion about the statistical inference on the jump θ. Section 4 is related to the estimation of variance σ^2.

2. F IS NORMAL WITH A KNOWN VARIANCE

In this section we consider the case that

$$F \sim N(0,\sigma^2) \tag{4}$$

and σ^2 is known. The method is based upon the following theorem.

THEOREM 1. Suppose that $X_1,\ldots,X_n$ are independently and identically distributed (iid) with a common distribution $N(a,\sigma^2)$. Let $\ell = \ell_n$ be a positive integer such that

$$\lim_{n\to\infty} \ell/n = 0, \qquad \lim_{n\to\infty} (\log n)^2/\ell = 0 \tag{5}$$

Define

$$Y_m = \frac{1}{\sqrt{2\ell}} \left(\sum_{m-\ell+1}^{m} X_i - \sum_{m-2\ell+1}^{m-\ell} X_i \right), \quad m = 2\ell,\ 2\ell+1,\ . \tag{6}$$

$$\xi_n = \max\{|Y_m| : m = 2\ell, 2\ell+1,\ldots,n\} \tag{7}$$

$$A_n(x) = [2\log(3n/2\ell-3)]^{-1/2}(x+2\log(3n/2\ell-3) + \tfrac{1}{2}\log\log(3n/2\ell-3) - \tfrac{1}{2}\log\pi) \tag{8}$$

Then, $$\lim_{n\to\infty} P(\xi_n/\sigma \le A_n(x)) = \exp(-2e^{-x}), \quad -\infty < x < \infty \tag{9}$$

Proof. Construct a Standard Brownian Motion $\{W(t),\ t \ge 0\}$ such that

$$W\left(\frac{3m}{2\ell}\right) = \sqrt{3/2\ell}\,(X_1+\ldots+X_m-ma)/\sigma, \quad m = 1,2,\ldots$$

and put

$$Z(t) = 3^{-1/2}[W(t)-2W(t+3/2) + W(t+3)], \quad t \ge 0$$

Then it is seen $Y_m = \sigma Z(3m/2\ell-3)$, $m = 2\ell,\ 2\ell+1,\ldots$. Put

$$\tilde{\xi}_n = \sup\{|Z(t)| : 0 \leq t \leq 3n/2\ell-3\}, \quad \eta_n = \tilde{\xi}_n - \xi_n/\sigma \tag{10}$$

We show that

$$\lim_{n\to\infty} \eta_n \sqrt{\log n} = 0, \quad \text{a.s.} \tag{11}$$

In fact, we have

$$0 \leq \eta_n \leq \frac{4}{\sqrt{3}} \sup\{|W(t+s) - W(t)| : \ 0 \leq s \leq 3/2\ell, \quad 0 \leq t \leq (3n-3)/2\ell-3\}$$

Insert $T = 3n/2\ell - 3$, $h = 3/2\ell$, $\varepsilon = 1$ and $\upsilon = (\sqrt{2}/4)\sqrt{\ell/\log n}\,\delta$ in lemma 1.2.1 of Csörgo and Revesz (1981), we get

$$P(\eta_n\sqrt{\log n} \geq \delta) \leq C\, n \exp(-\delta^2\ell/24\log n) \tag{12}$$

Here $\delta > 0$ is given and C is a constant not depending on n. Since $\ell/(\log n)^2 \to \infty$. (11) follows from (10) and Borel-Cantelli Lemma.

$Z(t)$ is a stationary Gaussian process, with $EZ(t) = 0$, and

$$P(\tau) = \mathrm{Cov}(Z(t), Z(t+\tau)) = \begin{cases} 1 - |\tau|, & 0 \leq |\tau| \leq 3/2 \\ -1 + |\tau|/3, & 3/2 < |\tau| \leq 3 \\ 0 & 3 \leq |\tau| < \infty \end{cases}$$

Thus, the conditions of a theorem of Qualls and Watanabe (1972) are met and we have

$$\lim_{n\to\infty} P(\tilde{\xi}_n \leq A_n(x)) = \exp(-2e^{-x}) \tag{13}$$

For n large we have

$$\begin{aligned} A_n(x+\Delta x) - A_n(x) &\geq \Delta x/\sqrt{2\log n}, \quad \Delta x > 0, \\ &\leq \Delta x/\sqrt{2\log n}, \quad \Delta x < 0. \end{aligned}$$

Hence

$$P(\tilde{\xi}_n \le A_n(x-|\Delta x|)) - P(\eta_n \ge |\Delta x|/\sqrt{2\log n})$$

$$\le P(\xi_n/\sigma \le A_n(x))$$

$$\le P(\tilde{\xi}_n \le A_n(X+|\Delta x|) + P(\eta_n \ge |\Delta x|/\sqrt{2\log n}) \tag{14}$$

For n large. From (12) - (14), letting $n \to \infty$ and then $\Delta x \to 0$, we get (9). Theorem 1 is proved.

Theorem 1 suggests a way to test the hypothesis (3): For the chosen size α, $0 < \alpha < 1$, solve the equation $\exp(-2e^{-x}) = 1 - \alpha$, the solution is

$$x(\alpha) = -\log(-\tfrac{1}{2}\log(1-\alpha))$$

Calculate $d = 2\ell/n$, $C_n(\alpha,d) = A_n(x(\alpha))$, and reject the null hypothesis (3) when and only when

$$\xi_n > \sigma\, C_n(\alpha,d) \tag{15}$$

Under condition (3), this test has an asymptotic size α as the sample size n tends to infinity.

We give an estimate of the power $\beta(\theta) = \beta_n(\theta,\sigma)$ of this test. For this purpose, let r be the integer satisfying

$$r/n \le t_0 < (r+1)/n \tag{16}$$

Then $Y_{r+\ell} \sim N(\sqrt{\ell/2}\,\theta,\sigma^2)$. Therefore,

$$\beta(\theta) \ge P(|Y_{r+\ell}| > \sigma\, C_n(\alpha,d))$$

$$= \Phi(\frac{|\theta|\sqrt{\ell}}{\sqrt{2}\,\sigma}\, C_n(\alpha,d)) + \Phi(-\frac{|\theta|\sqrt{\ell}}{\sqrt{2}\,\sigma} - C_n(\alpha,d))$$

$$> \Phi(\frac{|\theta|\sqrt{\ell}}{\sqrt{2}\,\sigma} - C_n(\alpha,d)) \tag{17}$$

where Φ is the distribution function of $N(0,1)$. From this inequality it is seen as is obvious intuitively, that for gaining a larger $\beta(\theta)$, one should give ℓ a larger value. (17) Also suggests, as is again obvious intuitively, that $\beta(\theta)$ increases steadily with $|\theta|/\sigma$, the magnitude of the jump measured against the standard deviation σ. Practically, the presence of a change-point is of importance only when the ratio $|\theta|/\sigma$ is substantially different from zero.

Now consider the interval estimation of the change-point t_0. The existence of t_0 may be a fact known in advance, but usually it is evidenced by the rejection of the null hypothesis (3). Regardness which is the case, we adopt the following

Rule. Find integer k such that $|Y_k| = \xi_n$. Take $[(k-2\ell)/n, k/n]$ as the confidence interval of t_0.

The length of this interval is $d = 2\ell/n$. Therefore, to enhance the accuracy of this estimate, one should choose a small ℓ. But there are other things to be considered. First, if the existence of t_0 is to be decided by the test defined above, a small ℓ increases the risk of false acception of the hypothesis (3), when in fact it is untrue. Second, a small ℓ corresponds a low confidence coefficient of the interval. To be more specific, we give an estimate of the confidence coefficient γ of this interval, as follows:

$$\gamma = P\left(\frac{k-2\ell}{n} \leq t_0 \leq \frac{k}{n}\right)$$

$$\geq P\left(\left\{\sup_{m\in(r,\gamma+2\ell)} |Y_m| \leq \sigma\, C_n(\alpha,d)\right\} \bigcap \left\{|Y_{r+0}| > \sigma\, C_n(\alpha,d)\right\}\right)$$

Since $X_m, X_{m-1}, \ldots, X_{m-2\ell-1}$ are iid when $m \in (r, r+2\ell)$, by Theorem 1 we have

$$P\left(\sup_{m\in(r,r+2\ell)} |Y_m| \leq \sigma\, C_n(\alpha,d)\right) \geq P\left(\xi_n \leq \sigma\, C_n(\alpha,d)\right) \doteq 1-\alpha \tag{18}$$

where the probability $P(\xi_n \leq \sigma\, C_n(\alpha,d))$ is computed under the condition that $X_1,\ldots,X_n$ are iid. Hence, neglecting the error in the latter equality of (18), we have

$$\gamma \geq (1-\alpha) + P(|Y_{r+\ell}| \geq \sigma\, C_n(\alpha,d)) - 1$$

$$\geq \Phi\left(\frac{|\theta|\sqrt{\ell}}{\sqrt{2}\,\sigma} - C_n(\alpha,d)\right) - \alpha \tag{19}$$

and this inequality suggests that γ should increase with $|\theta|/\sigma$.

Using (19), we can give the following important question an approximate solution: Form a confidence interval of t_0 with prescribed length d_0 and confidence coefficient $1-\alpha_0$. The question is how to choose ℓ and n. To do this, give α in (19) the value $\alpha_0/2$, and solve the equation

$$\Phi(|\theta|\sqrt{\ell}/(\sqrt{2}\,\sigma) - C_n(\alpha,d)) = 1 - \alpha_0/2$$

to obtain

$$\ell \doteq 2\left(\frac{|\theta|}{\sigma}\right)^{-2}\left(C_n\left(\frac{\alpha_0}{2},d_0\right) + U_{\alpha_0/2}\right)^2, \quad n \doteq 2\ell/d_0\ . \tag{20}$$

Here $U_{\alpha_0/2}$ is the upper $\alpha_0/2$ - point of $N(0,1)$.

The solution (20) has the trouble that it involves the parameters θ and δ, the former is surely unknown and the latter is usually unknown. In some cases it may be feasible to take some preliminary samples to give a crude estimation of them, but we recommend the following procedure: Decide by practical consideration a constant M such that only when $|\theta|/\sigma \geq M$, the change-point t_0 is of any real importance. Replace $|\theta|/\sigma$ in (20) by this M.

For example, take $\alpha_0 = 0.05$, $d_0 = 0.1$, $M = 2$, (20) gives

$$\ell = 19.46, \qquad n = 389.2$$

The result seems rather good as it is comparable with the sample size which is needed in estimating the probability p of the binomial $B(n,p)$. The sample size needed to guarrantee a confidence interval of p not longer than 0.1 and with a confidence coefficient not smaller than 0.95 is roughly 384. So when $|\theta|/\sigma$ is not smaller than 2, which seems a moderate requirement in practice, we are in a situation comparable to one seeking an estimation of the binomial p.

3. F IS NORMAL WITH VARIANCE UNKNOWN

When the variance σ^2 is unknown, we use the sample $X_1, \ldots, X_n$ to estimate it. Denote by $\hat{\sigma}^2$ the estimator of σ^2, we use $\hat{\sigma}$ to replace σ in (15) to perform the test. In order that the resulted test still has an asymptotic size α, the estimator $\hat{\sigma}^2$ must satisfy certain condition stated in the following theorem.

THEOREM 2. Suppose that the conditions of Theorem 1 are met, and $\hat{\sigma}^2$ is an estimator of σ^2 satisfying

$$\lim_{n\to\infty} \left| \hat{\sigma}^2 - \sigma^2 \right| \log n \overset{p}{=} 0 \tag{21}$$

$\underline{p}$ means convergence in probability. Then

$$\lim_{n\to\infty} P(\xi_n/\hat{\sigma} \leq A_n(x)) = exp(-2e^{-x}) \tag{22}$$

Proof follows easily from Theorem 1. For, from (21) we have $|\hat{\sigma} - \sigma| \log n \xrightarrow{p} 0$. Thus, for arbitrarily given $\varepsilon > 0$ we have

$$P(\xi_n/\sigma \leq A_n(x)(1-\varepsilon/\log n)) - P(|\hat{\sigma} - \sigma| \log n \geq \varepsilon)$$

$$\leq P(\tilde{\xi}_n/\hat{\sigma} \leq A_n(x))$$

$$\leq P(\xi_n/\sigma \leq A_n(x)(1+\varepsilon/\log n)) + P(|\hat{\sigma} - \sigma| \log n \geq \varepsilon)$$

Given $\delta > 0$. For n large we have

$$A_n(x-\delta) \leq A_n(x)(1-\varepsilon/\log n) \leq A_n(x)(1+\varepsilon/\log n) \leq A_n(x+\delta)$$

Therefore, for n large we have

$$P(\xi_n/\sigma \leq A_n(x-\delta)) - P(|\hat{\sigma} - \sigma|\log n \geq \varepsilon)$$

$$\leq P(\xi_n/\hat{\sigma} \leq A_n(x))$$

$$\leq P(\xi_n/\sigma \leq A_n(x+\delta)) + P(|\hat{\sigma} - \sigma|\log n \geq \varepsilon)$$

Letting $n \to \infty$ and then $\delta \to 0$, we obtain (21)

Estimation of the variance σ^2.

Under the normal assumption here, the maximum likelihood estimator has the form (26) given below. The following theorem shows that this estimator satisfies (21).

THEOREM 3. Suppose that $\{N_1, N_2, \ldots\}$ is a sequence of positive integers tending to infinity. For each n, there are given N_n independent variables X_{ni}, $i = 1,\ldots,N_n$, such that

$$X_{ni} \sim F(x), \quad i = 1,\ldots,m_n, \quad X_{ni} \sim F(x-\theta), \quad i = m_n+1,\ldots,N_n$$

where F is a distribution function whose moment of $(2+\delta)$-th order exists, θ is a constant. Define

$$\overline{X}_{n1c} = \sum_1^c X_{ni}/C, \quad \overline{X}_{n2c} = \sum_{c+1}^{N_n} X_{ni}/(N_n-c), \quad \overline{X}_n = \sum_1^{N_n} X_{ni}/N_n \tag{23}$$

$$S_{nc}^2 = \sum_1^c (X_{ni}-\overline{X}_{n1c})^2 + \sum_{c+1}^{N_n} (X_{ni}-\overline{X}_{n2c})^2, \quad c = 1,\ldots,N_n \tag{24}$$

$$S^2(n) = \min\left\{S_{nc}^2 : C = 1,\ldots,N_n\right\} \tag{25}$$

$$\hat{\sigma}_n^2 = \hat{\sigma}^2 = S^2(n)/N_n \tag{26}$$

Then we have,

$$\lim_{n\to\infty} |\hat{\sigma}_n^2 - \sigma^2| \log N_n \overset{P}{=} 0 \tag{27}$$

where σ^2 is the variance of F.

Proof. For simplicity of writing we shall in the sequel omit the symbol n in all notations. Thus, X_{ni}, m_n, N_n, $\overline{X}_{n1c}$, $\overline{X}_n$, S_{nc}^2, $S^2(n)$ and $\hat{\sigma}_n^2$ will be abbreviated to X_i, m, N, $\overline{X}_{1c}$, $\overline{X}$, S_c^2, S^2 and $\hat{\sigma}^2$ respectively. Put $Y_i = X_i - EX_i$, $T_i = Y_1 + \ldots + Y_i$. Given $\varepsilon > 0$, from Kolmogorov inequality we have

$$P(|T_c/C| \le \varepsilon/\log N, \quad N\varepsilon/\log N \le c \le N) \ge 1 - D(\log N)^4/N \tag{28}$$

$$P(|(T_N-T_c)/(N-c)| \le \varepsilon/\log N, \quad 1 \le c \le N(1-\varepsilon/\log N) \ge 1 - D(\log N)^4/N \tag{29}$$

Here and in the following we shall use D to denote a constant not depending on n, which may assume different values in each of its appearance. From (28), (29), we have

$$P(|T_c/c| \le \varepsilon/\log N, \quad |(T_N-T_c)/N-c)| \le \varepsilon/\log N :$$

$$N\varepsilon/\log N \le c \le N(1-\varepsilon/\log N)) \ge 1 - D(\log N)^4/N. \tag{30}$$

First consider the case $\theta = 0$. Since

$$S_c^2 - S_N^2 = -\frac{C(N-c)}{N} (\overline{X}_{1c}-\overline{X}_{2c})^2 \tag{31}$$

and $\overline{X}_{1c} - \overline{X}_{2c} = T_c/C - (T_N-T_c)/(N-c)$ when $\theta = 0$, We have by (30) and (31) that

$$P(S_c^2 - S_N^2 \geq -\, N\varepsilon^2/(\log N)^2, \quad N\varepsilon/\log N \leq c \leq N(1-\varepsilon/\log N))$$

$$\geq 1 - D(\log N)^4/N \,. \tag{32}$$

For $C \leq N\varepsilon/\log N$ we have

$$S_c^2 - S_N^2 \geq -\sum_{1}^{N\varepsilon/\log N} (X_1 - \overline{X}_{2c})^2$$

$$\geq -\,2\Big(\sum_{1}^{N\varepsilon/\log N} X_1^2 + \frac{\varepsilon}{\log N}\Big(1 - \frac{\varepsilon}{\log N}\Big)^{-1} \sum_1^N X_i^2\Big) \tag{33}$$

For $C \geq N(1-\varepsilon/\log N)$ one obtains similarily

$$S_c^2 - S_N^2 \quad -\,2\Big(\sum_{N(1-\varepsilon/\log N)}^{N} X_i^2 + \frac{\varepsilon}{\log N}\Big(1 - \frac{\varepsilon}{\log N}\Big)^{-1} \sum_1^N X_i^2\Big) \tag{34}$$

From (32) - (34), and putting

$$Q = \frac{\varepsilon^2}{\log N} + \frac{2\varepsilon}{1-\varepsilon}\,\frac{1}{N}\sum_1^N X_i^2 + 2\varepsilon\,\frac{\log N}{N\varepsilon}\Big(\sum_1^{N\varepsilon/\log N} X_i^2 + \sum_{N(1-\varepsilon/\log N)}^{N} X_i^2\Big) \tag{35}$$

We have

$$P(|\hat{\sigma}^2 - S_N^2/N|\log N \geq Q) \leq D(\log N)^4/N \tag{36}$$

Denoting by a the expectation of variable X_1, we have

$$\lim_{n\to\infty} P\Big(Q \leq \frac{6\varepsilon}{1-\varepsilon}(\sigma^2 + \sigma^2 + 1)\Big) = 0 \tag{37}$$

Further, it can easily be shown that under the conditions of the present theorem, there exists $S' > 0$ such that

$$\lim_{n\to\infty} N^{S'}(\sigma^2 - S_N^2/N) \overset{P}{=} 0 \tag{38}$$

From (36) - (38), (27) follows.

Now turn to the case of $\theta \neq 0$. Put

$$\xi = \xi_{nmc} = \begin{cases} \sum_{m+1}^{c} x_i/(c-m), & c < m \\ \sum_{c+1}^{m} x_i/(m-c), & c < m \end{cases}$$

Then one can easily verify that

$$S_i^2 - S_m^2 \geq \begin{cases} -(c-m)(\overline{X}_{2c}-\xi)^2, & c > m \\ -(m-c)(\overline{X}_{1c}-\xi)^2, & c < m \end{cases}$$

Given $\varepsilon > 0$ small enough. Write

$$J = \{c : N\varepsilon/\log N \leq c \leq N(1-\varepsilon/\log N)$$

and separate the following four cases:

1. $C \bar{\in} J$.

As in the case of $\theta = 0$, we have (33) and (34)

2. $C \in J, \quad N^{1/3} \leq |c-m| \leq N^{7/12}$

By Kolmogorov inequality, we have

$$P(\ \overline{X}_{2c}-(a+\theta)| \leq \varepsilon/\log N, \quad C \in J) \geq 1 - D(\log N)^4/N \tag{40}$$

$$P(|\xi-(a+\theta)| \leq \varepsilon/\log N, \ N^{1/3} \leq c - m \leq N^{7/12}) \geq 1 - D(\log N)^2 N^{-1/12} \tag{41}$$

$$P(|\xi-a| \leq \varepsilon/\log N, \quad N^{1/3} \leq m - c \leq N^{7/12}) \geq 1 - D(\log N)^2 N^{-1/12} \tag{42}$$

From (40), (41), and the first part of (39), we get

$$P(S_c^2-S_m^2 \geq - 4\varepsilon^2 N^{7/12}/(\log N)^2, \ c \in J, \ N^{1/3} \leq c - m \leq N^{7/12})$$

$$\geq 1 - D(\log N)^2 N^{-1/12} \tag{43}$$

The case of c < m can be handled by (40) (42), and the second part of (39), resulting in an inequality similar to (43). Combining the two, we obtain

$$P(S_c^2-S_m^2 \geq -4\varepsilon^2N^{7/12}/(\log N)^2,\ c \in J,\ N^{1/3} \leq |c-m| \leq N^{7/12})$$

$$\geq 1 - D(\log N)^2N^{-1/12} \tag{44}$$

3. $C \in J$, $|c-m| \geq N^{7/12}$

Kolmogorov inequality gives

$$P(|\xi-(a+\theta)| \leq \varepsilon/\log N,\ c - m \geq N^{7/12}) \geq 1 - D(\log N)^2N^{-1/6} \tag{45}$$

$$P(|\xi-a| \leq \varepsilon/\log N,\ m - c \geq N^{7/12}) \geq 1 - D(\log N)^2N^{-1/6} \tag{46}$$

From (40), (45) and the first part of (39), we get an inequality similar to (43). Likewisely, from (40), (46) and the second part of (39), we get another inequality. Combine those two, we get

$$P(S_c^2-S_m^2 \geq -4\varepsilon^2N/(\log N)^2,\ c\in J,\ |c-m| \geq N^{7/12})$$

$$\geq 1 - D(\log N)^2N^{-1/6} \tag{47}$$

4. $c\in J$, $|c-m| \leq N^{1/3}$

For c m, we have by the first part of (39)

$$S_c^2 - S_m^2 \geq -2N^{1/3}(\overline{X}_{2c}^2+\xi^2) \geq -2N^{1/3}(\overline{X}_{2c}^2+\sum_m^{m+N^{1/3}} x_i^2) \tag{48}$$

From (40) we have $P(\overline{X}_{2c}| \leq |a|+|\theta|+1,\ c \in J) \geq 1 - D(\log N)^3/N$.

Also it is easy to see that

$$P\left(\sum_{m}^{m+N^{1/3}} x_i^2 \leq N^{1/2}\right) \geq 1 - DN^{-1/6}$$

Hence, from (48) we get

$$P(S_c^2 - S_m^2 \geq -N^{6/7}, \quad c\in J, \quad m \leq c \leq m + N^{1/3}) \geq 1 - DN^{-1/6}$$

This combining with the similar inequality obtained for the case of $c < m$, gives

$$P(S_c^2 - S_m^2 \geq -N^{6/7}, \quad c\in J, \quad |c-m| \leq N^{1/3}) \geq 1 - DN^{-1/6}$$

Combine this, (33), (34), (44), (47), define Q as before, and notice that $S_N^2 \geq S_m^2 \geq S^2$, we obtain

$$P(|S_m^2/N-\sigma\ |\log N \geq Q) \leq D(\log N)^2 N^{-1/12} \tag{49}$$

Similar to the case of $\theta = 0$, here we can still prove that there exists a constant R not depending on n, such that $\lim_{n\to\infty} P(Q \geq R\varepsilon) = 0$. Also, under the conditions of the present theorem, we can find $\delta' > 0$ such that $N^{\delta'}|\sigma^2 - S_m^2/N| \to 0$ in probability. These facts, combined with (49), give (27), and the theorem is proved.

4. BIAS OF THE VARIANCE ESTIMATOR

Maintain the notations of section 3. Since $S^2 = \min\{S_c^2\} \leq S_m^2$, and $ES_m^2/(N-2) = \sigma^2$, we see that even if we replace $\hat{\sigma}^2$ by $\tilde{\sigma}^2 = S^2/(N-2)$, $\tilde{\sigma}^2$ still underestimates σ^2. Intuitively it seems clear that the bias should be more serious if $|\theta|/\sigma$ is small. So we propose to look in some detail the bias in case that $\theta = 0$.

From (31) we have

$$\frac{1}{2}(S_N^2-S^2) = \max_{1\leq c\leq N} \frac{N}{C(N-c)}(T_c-\frac{c}{N}T_N)^2 = \left\{\sup_{0<t<1} |Z_N(t)|/\sqrt{t(1-t)}\right\}^2 \triangleq \eta_n$$

where $T_c = \sum_1^c (X_i-a)/\sigma$, and

$$Z_N(t) = \begin{cases} (T_{[(N+1)t]} - [(N+1)t]T_N/N)/\sqrt{N}, & 0 \leq t < 1 \\ 0 \quad , & t = 1 \end{cases}$$

When X_1 is normally distributed, or more generally, when $E|X_1|^\delta < \infty$ for some $\delta > 2$, the following asymptotic distribution is valid (Yao and David (1984), Csorgo and Horvath (1986)):

$$\lim_{n\to\infty} P(\sqrt{2\log_2 n}\,\sqrt{\eta_n} - (2\log_2 n + \tfrac{1}{2}\log_3 n - \tfrac{1}{2}\log\pi) < x) = \exp(-2e^{-x})$$

Here $\log_{k+1}(x) = \log\log_k(x)$, $\log_1 x = \log x$. From this, we deduce the following asymptotic distribution for the bias

$$\lim_{n\to\infty} P(\frac{S_N^2-S^2}{\sigma^2} < 2\log_2 N + \log_3 N - \log\pi + x) = \exp(-2e^{-x/2}) \tag{50}$$

This asymptotic result suggests that the bias has a form

$$S_N^2 - S^2 = 2\log_2 N + \log_3 N + Q_p(1)$$

Further, since the distribution function $\exp(-2e^{-x/2})$ has an expectation $2\gamma + 2\log 2 = 2.5 +$ (γ - the Euler constant), (50) suggests that when N is large, we may take

$$S^2/(N-1-(2\log_2 N+\log_3 N-\log\pi) - 2(\gamma+\log 2)) = S^2/(N-2\log_2 N-\log_3 N-2.4)$$

as the estimator of σ^2, in order to bring down the bias.

In some applications we might know in advance that there exist some constants λ_1, λ_2, $0 < \lambda_1 < \lambda_2 < 1$, such that the change-point t_0 lies in the interval $[\lambda_1,\lambda_2]$. In this case we may use

$$S^2(\lambda_1,\lambda_2) = \min(S_c^2 : \lambda_1 N \leq c \leq \lambda_2 N)$$

instead of S^2. We have

$$\frac{1}{\sigma^2}(S_N^2-S^2(\lambda_1,\lambda_2)) = \sup_{\lambda_1\leq t\leq\lambda_2} |Z_N(t)|/\sqrt{t(1-t)} \triangleq v_N$$

It is shown (Csorgo and Horvath (1986)) that as $N \to \infty$, v_N converges to $v = \sup(|V(s)| : 0 \leq S \leq \frac{1}{2}\log(\lambda_2(1-\lambda_1)/\lambda_1(1-\lambda_2)))$, where $\{V(s) : S \geq 0\}$ is the Ornstein - Uhlenbeck process, i.e., a Gaussian process with mean zero and covariance function $\exp(-|t-s|)$. The distribution function of v has been tabulated by DeLong (1981). Put $E(v) = C(\lambda_1,\lambda_2)$, it is reasonable to take $S^2(\lambda_1,\lambda_2)/(N-1-C(\lambda_1,\lambda_2))$ as an estimator of σ^2.

An upper bound of $C(\lambda_1,\lambda_2)$ can be estimated in the following way. Put $\lambda = \min(\lambda_1, 1-\lambda_2)$. We have

$$\frac{1}{\sigma^2}(S_N^2 - S^2(\lambda_1,\lambda_2)) \le \max\left\{\frac{N}{c(N-c)}(T_c - \frac{c}{N}\,T_N)^2 \;:\; \lambda N \le c \le (1-\lambda)N\right\}$$

$$\le \frac{1}{\lambda(1-\lambda)}\left\{\max_{\lambda N \le c \le (1-\lambda)N}\left|\frac{1}{\sqrt{N}}\,T_c - \frac{1}{\sqrt{N}}\,\frac{c}{N}\,T_N\right|\right\}^2$$

$$\le \frac{1}{\lambda(1-\lambda)}\left\{\max_{\lambda N \le c \le (1-\lambda)N}|B(t)|\right\}^2 \le \frac{1}{\lambda(1-\lambda)}\left\{\max_{0 \le t \le 1}|B(t)|\right\}^2$$

Where $\{B(t),\quad 0 \le t \le 1\}$ is the Standard Brownian Bridge.

The distribution function of $\max_{0 \le t \le 1}|B(t)|$ is well-known as $1 - \sum_1^\infty 2(-1)^{k+1}\exp(-2k^2x^2)$. From which it is easy to get

$$E\left\{\max_{0 \le t \le 1}|B(t)|\right\}^2 = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \ldots \approx 0.85$$

Hence,

$$C(\lambda_1,\lambda_2) \le 0.85\ \lambda^{-1}(1-\lambda)^{-1}$$

Therefore, if we use

$$\hat{\sigma}^2(\lambda_1,\lambda_2) = S^2(\lambda_1,\lambda_2)/(N-2)$$

to estimate σ^2, the relative bias does not exceed

$$[\sigma^2 - E\hat{\sigma}^2(\lambda_1,\lambda_2)]/\sigma^2 = (0.85\ \lambda^{-1}(1-\lambda)^{-1} - 1)/(N-2) \qquad (51)$$

For example, take $\lambda_1 = 1 - \lambda_2 = 0.1$. This value can perhaps be considered as small enough for many applications. Take $N = 100$. According to (51), the relative bias does not exceed 8.6%, which is reasonably good, considering the relatively small sample size 100 in estimating the variance of a complicated model.

5. F IS NON-NORMAL

When the distribution of the random error $e(t)$ is non-normal, we can use the theory of strong approximation of partial sums of iid. variables by Brownian Motion Process to give extensions of Theorem 1 to non-normal cases. In this way the methods of previous sections can still be applied.

THEOREM 4. Replacing the assumption of Theorem 1 that $X_1 \sim N(a,\sigma^2)$ by

$$E(\exp(tX_1)) < \infty \quad \text{for } |t| \text{ small enough} \tag{52}$$

Then the conclusion of Theorem 1 remains valid.

Proof. Put

$$S_k = (X_1 + \dots + X_k - ka)/\sigma, \quad k = 1,2,\dots$$

According to Komlós and others (1975, 1976), there exists a Brownian Motion process $\{W(t),\ t \geq 0\}$ such that

$$\limsup_{n\to\infty} \sup_{k\leq n} |S_k - W(k)|/\log n < \infty, \quad \text{a.s.} \tag{53}$$

Since

$$Y_m/\sigma = [(S_m - S_{m-\ell}) - (S_{m-\ell} - S_{m-2\ell})]/\sqrt{2\ell}$$

We have for $m \leq n$

$$|Y_m/\sigma - [W(m) - 2W(m-\ell) + W(m-2\ell)]/\sqrt{2\ell}| \leq 4 \sup_{1\leq k\leq n} |S_k - W(k)|/\sqrt{2\ell} \tag{54}$$

According to (53), noticing that $\log n/\sqrt{\ell} \to 0$ as $n \to \infty$, we get

$$\lim_{n\to\infty}\left[\sup_{m\le n}|Y_m/\sigma-[W(m)-2W(m-\ell)+W(m-2\ell)]/\sqrt{2\ell}|\right]=0, \quad \text{a.s.} \tag{55}$$

From Theorem 1, we have

$$\lim_{n\to\infty} P\{\sup([W(m)-2W(m-\ell)+W(m-2\ell)]/\sqrt{2\ell} : m = 2\ell,\ldots,n)\}$$

$$\le A_n(x) = \exp(-2e^{-x}), \quad -\infty < x < \infty \tag{56}$$

Finally, (7) follows from (55) and (56). Theorem 4 is proved.

Since under the assumption (52) the conclusion of Theorem 3 is also true, it follows that the method of the previous two sections applies to the case in which (52) is true.

The condition ensuring the asymptotic results in sections 2, 3, can further be substantially weakened.

THEOREM 5. Replacing the assumptions of Theorem 1 that $X_1 \sim N(a,\sigma^2)$ and $(\log n)^2/\ell \to 0$ by : For some $\delta > 2$

$$E|X_1|^\delta < \infty \tag{57}$$

$$\lim_{n\to\infty} n^{2/\delta}/\ell = 0 \tag{58}$$

Then the conclusion of theorem 1 remains valid.

The proof parallels that of Theorem 4, with the help of another result of Kimlos and others, which asserts

$$\lim_{n\to\infty}\left\{\sup_{k\le n}|S_k-W(k)|n^{-1/\delta}\right\}=0, \quad \text{a.s.}$$

under condition (57).

6. ESTIMATION OF JUMP θ

To form a point estimate of the jump θ at the change point t_0, we suggest the following procedure:

1. Find k such that $|Y_k| = \xi_n$.

Compute

$$\hat{\theta} = \hat{\theta}_n = \frac{1}{n-k} \sum_{k+1}^{n} X_i - \frac{1}{k-2\ell} \sum_{1}^{k-2\ell} X_i \tag{59}$$

$\hat{\theta}$ is taken as an estimate of θ.

When $k = 2\ell$, or $k = n$, $\hat{\theta}$ is not defined. In general, if k is too near 2ℓ or n, this would imply that the change-point t_0 is too near 0 or 1, and the samples at our disposal are perhaps not enough to give a reasonable estimate.

For an interval estimation of θ, we prove the following asymptotic theorem about $\hat{\theta}$.

THEOREM 6. Suppose that the δ-th order moment of distribution F is finite for some $\delta > 2$. Further, $\theta \neq 0$ and ℓ satisfies (58) and $\ell/n \to 0$. Then as $n \to \infty$ we have

$$(\sqrt{nt_0(1-t_0)}/\sigma)(\hat{\theta}-\theta) \xrightarrow{L} N(0,1) \tag{60}$$

where $\xrightarrow{L}$ stands for convergence in law.

Proof. Without losing generality, we assume $a = 0$, $\sigma = 1$ (see section 1). Using Theorem 5 and slightly modifying the argument of section 2 we easily show that

$$\lim_{n\to\infty} P(nt_0 \le k \le nt_0 + 2\ell) = 1 \tag{61}$$

Define $n_1 = [nt_0 - 2\ell]$, $n_2 = n - [nt_0 + 2\ell] - 1$, and

$$T_n = \sqrt{nt_0(1-t_0)}\left(\frac{1}{n_2}\sum_{n-n_2+1}^{n}(X_i - \theta) - \frac{1}{n_1}\sum_{1}^{n_1} X_i\right) \tag{62}$$

Since $\ell/n \to 0$, $0 < t_0 < 1$, we have as $n \to \infty$

$$T_n \xrightarrow{L} N(0,1) \tag{63}$$

Now,

$$T_n - \sqrt{nt_0(1-t_0)}\,(\hat{\theta}-\theta) = \sqrt{nt_0(1-t_0)}\left\{\frac{n-k-n_2}{n_2(n-k)}\sum_{n-n_2+1}^{n}(X_i-\theta)\right.$$

$$\left. - \frac{1}{n-k}\sum_{k+1}^{n-n_2}(X_i-\theta) - \frac{k-2\ell-n_1}{n_1(k-2\ell)}\sum_{1}^{n_1} X_i + \frac{1}{k-2\ell}\sum_{1}^{k-2\ell} X_i\right\}$$

Therefore,

$$\mathrm{Sup}\{|T_n - \sqrt{nt_0(1-t_0)}(\hat{\theta}-\theta)| : nt_0 \le k \le nt_0 + 2\ell\}$$

$$\le C\left\{n^{-3/2}\ell.\mathrm{Sup}(|\sum_{j}^{n}(X_i-\theta)| : j \ge n - n_2 + 1)\right.$$

$$+ n^{-1/2}\mathrm{Sup}(|\sum_{j}^{n-n_2}(X_i-\theta)| : [nt_0] + 1 \le j \le n - n_2)$$

$$+ n^{-3/2}\ \ell.\mathrm{Sup}(|\textstyle\sum_1^j X_i| : 1 \le j \le n_1)$$

$$\left. + n^{-1/2}\mathrm{Sup}(|\textstyle\sum_{n_1+1}^{j} X_i| : n_1 + 1 \le j \le [nt_0])\right\} \triangleq \sum_{i=1}^{4} I_{ni} \tag{64}$$

where C is a constant not depending on n. Since $\ell/n \to 0$, $X_1, \ldots, X_n$ are independent with variance 1, $EX_i = \theta$ for $i \geq n - n_2 + 1$, it follows from Dosker's Theorem that

$$I_{n1} \xrightarrow{P} 0, \quad \text{as} \quad n \to \infty \tag{65}$$

Similarily,

$$I_{n3} \xrightarrow{P} 0, \quad \text{as} \quad n \to \infty \tag{66}$$

Now define $\tilde{X}_i = X_i + \theta$ for $1 \leq i \leq [nt_0]$, $\tilde{X}_i = X_i$ for $[nt_0] + 1 \leq i \leq n$. Given arbitrarily $\varepsilon > 0$, we have by Donsker's Theorem

$$I_{n2} \leq n^{-1/2} \operatorname{Sup}(|\sum_{j}^{n-n_2} (\tilde{X}_i - \theta) \quad : [n(t_0 - \varepsilon)] \leq j \leq n - n_2)$$

$$\xrightarrow{L} \operatorname{Sup}(|W(t)| : t_0 - \varepsilon \leq t \leq t_0), \quad \text{as} \quad n \to \infty$$

$$\longrightarrow \quad 0, \quad \text{a.s.} \quad , \quad \text{as} \quad \varepsilon \to 0$$

Hence,

$$I_{n2} \xrightarrow{P} 0, \quad \text{as} \quad n \to \infty \tag{67}$$

Similarly

$$I_{n4} \xrightarrow{P} 0, \quad \text{as} \quad n \to \infty \tag{68}$$

Summing up (65) - (68), we get

$$\operatorname{Sup}(|T_n - \sqrt{nt_0(1-t_0)}(\hat{\theta} - \theta)| : nt_0 \leq k \leq nt_0 + 2\ell) \xrightarrow{P} 0$$

as $n \to \infty$. From this and (61), (63), we obtain (60). The proof is concluded.

It is easy to see that $\hat{t}_0 = (k-\ell)/n$ is a consistent estimator of t_0 (of course, when $\theta \neq 0$ and thus t_0 is well-defined). Earlier we have introduced a consistent estimate $\hat{\sigma}$ of σ. Substituting $\hat{t}_0$ for t_0 and $\hat{\sigma}$ for σ. We get the following theorem:

THEOREM 7. Suppose that the conditions of Theorem 6 are met, we have

$$(\sqrt{n\hat{t}_0(1-\hat{t}_0)}/\hat{\sigma})(\hat{\theta}-\theta) \xrightarrow{L} N(0,1) \tag{69}$$

as $n \to \infty$

When $\theta = 0$, in which case t_0 has no meaning, the statistic $\hat{t}_0$ is still well-defined. It is not known whether or not (69) is true for $\theta = 0$. So Theorem 7 cannot be employed to make tests for the hypothesis $\theta = 0$, but (69) can be used to form a confidence interval of θ, when $\theta \neq 0$ is assumed in advance, or as a result of the rejection of null hypothesis $\theta = 0$.

Of course, if X_1 is normally distributed, or more generally, X_1 satisfies condition (52), then the condition (58) in Theorem 6 and 7 can be weakened to $(\log n)^2/\ell \to 0$.

REFERENCES

[1] CSORGO, M., and HORVATH, L. (1986). Nonparametric Methods for Change Point Problems to Appear in *Handbook of Statistics* Vol. 7.

[2] CSORGO, M., and REVESZ, P. (1981). Strong Approximations in Probability and Statistics. Akademiai Kiado, Budapest.

[3] DeLONG, D. M. (1981). Crossing Probabilities for a Square Root Boundary by a Bessel Process. *Comm. Statist. A-Theor. Methods* A10 2197-2133.

[4] KOMLOS, J., MAJOR, P. and TUSNADY, G. (1975). An Approximation of Partial Sums of Independent R.V.'s and the Sample DF. I. *Z Wahrsche. Verw. Gebiete,* **32**, 111-131; II, (1976) **34**, 33-58.

[5] KRISHNAIAH, P. R., MIAO, B. (1986). Review About Estimates of Change Point, to Appear in *Handbook of Statistics*, Vol. 7.

[6] QUALLS, C., WATANABE, H. (1972). Asymptotic Properties of Gaussian Processes, *Ann. Math. Statist.*, **43**, 580-596.

[7] YAO, Y. C., and DAVID, R. A. (1984). The Asymptotic Behavior of the likelihood ratio statistic for testing a shift in mean in sequence of independent normal variates. Center for Stochastic Processes, Dept. of Statistics, University of North Carolina, Technical Report No.80.

[8] YIN, Y. Q., (1986). Detection of the Number, Locations and Magnitudes of Jumps, Technical Report.

END

DATE

FILMED

DEC.

1987